

FIG. 1A

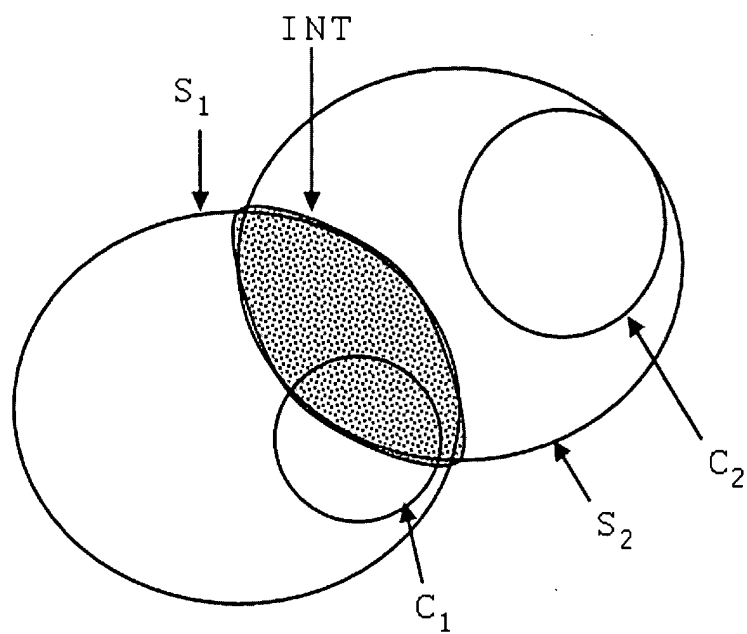


FIG. 1B

Start with concept C1 (210)
Perform a search for relevant documents to C1 (220)
Analyze the relevant documents for C1, and determine "expansion concepts" E1-1,E1-2,E1-3, etc... (230)
Perform a search for the expansion concepts for C1 - define the set S1 as the union of all results found from 220 and 230 (240)
For each document in S1 assign a relevance score r with respect to the original concept C1 (250)
Repeat Steps 10-50, replacing C1 with concept C2, and replacing set S1 with set S2 (260)
Find the intersection INT of the two sets S1 and S2 (270)
For each document d in the intersection INT compute: -- Relevance: $d_r = \text{relevance}(S1) * \text{relevance}(S2)$ -- Explanation of d is $\text{Expl}(d) = \text{the "concepts" that are contained in that document from the larger concept set CS. CS} = \{C1, C2, E1-1, E1-2, \dots, E1-n, E2-1, E2-2, \dots, E2-n\}$ (280)
Determine relationship between C1 and C2 based on relevance of documents (290)

FIG. 2

perform off-line (training) 304
perform relevance determination/ query expansion on-line 306

FIG. 3A

Submit initial user query to data source (310)
Retrieve and score initial result set - assigning a relevance score to each document (320)
Define the "positive set" to be those documents scoring over a specified relevance threshold (330)
Build a "positive set histogram" from the positive set (340)
Apply thresholding to remove features (350)
For the remaining features, rank them by expected entropy loss (360)
Choose the top ranked feature set as ranked in 360 (370)
Choose the expanded feature set by applying "concept constraints" (380)
Define the feedback-scoring-function (390)
Submit an OR query of the expanded features found from 380 (400)
Score returned documents by function in 390 (410)

FIG. 3B

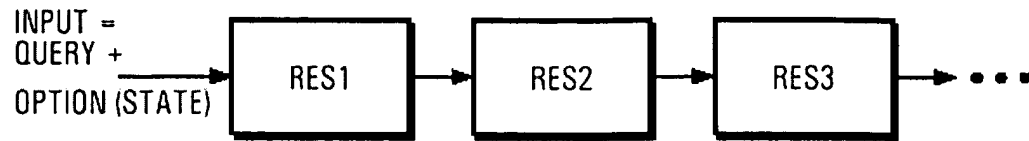


FIG. 4A

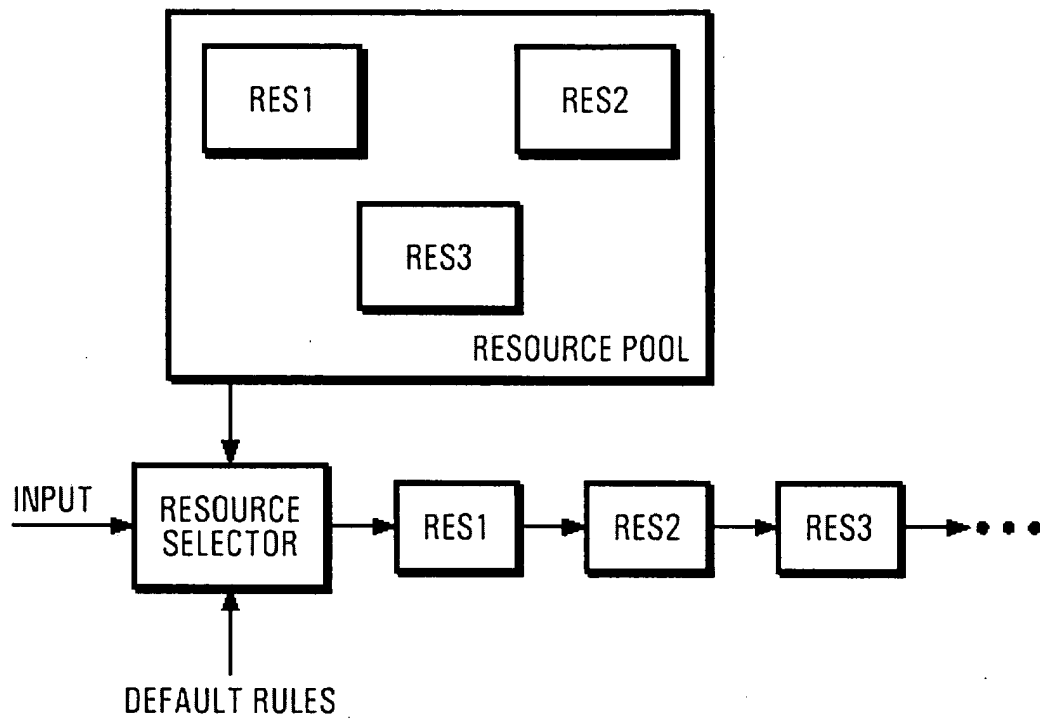


FIG. 4B

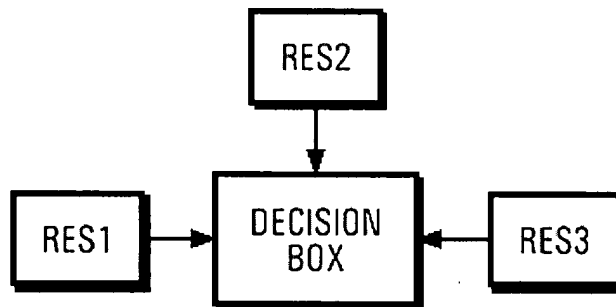


FIG. 4C

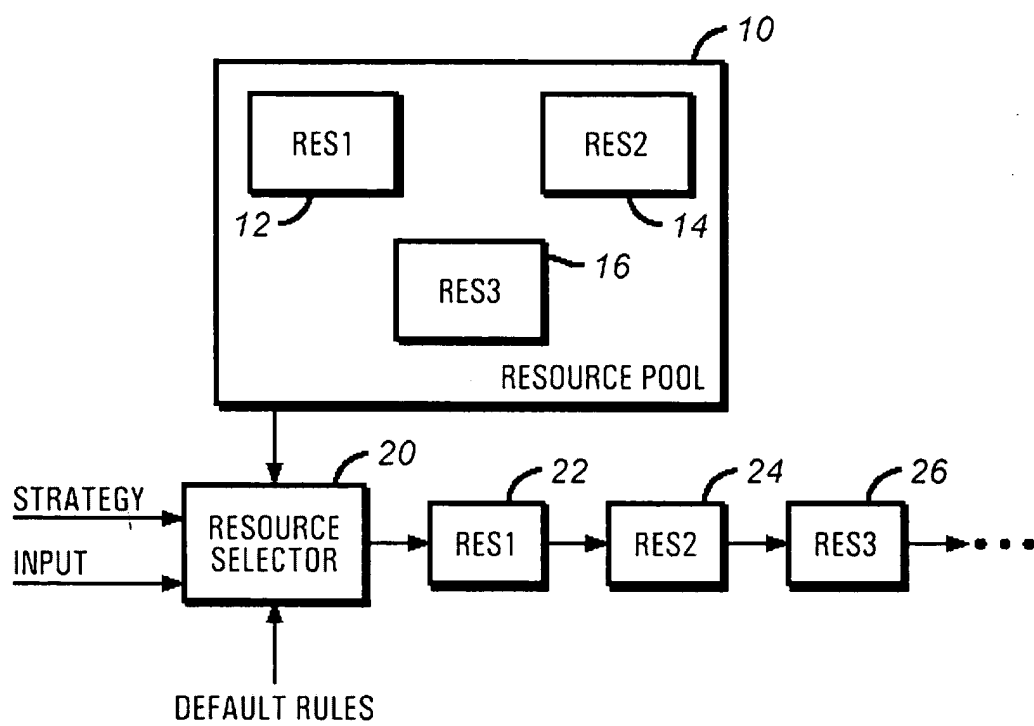


FIG. 4D